Person identification from walking sound on wooden floor

Downloaded from: https://research.chalmers.se, 2023-05-04 22:14 UTC

Citation for the original published paper (version of record):
Diapoulis, G., Rosas Perez, C., Larsson, K. et al (2018). Person identification from walking sound on wooden floor. Euronoise 2018: 1727-1732

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Person identification from walking sound on wooden floor

Georgios Diapoulis
Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.

Carmen Rosas
Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.

Krister Larsson
Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.

Wolfgang Kropp
Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.

Summary

Human activities in building structures may vary to a great extend and daily mobility may be the cause of noise and vibrations. We focused on walking sound to identify different individuals based on airborne sound recordings. Our aim was to provide a description of high-level acoustical features that corresponds to walking sound and person identification. We present two levels of abstraction. The first level builds upon principal component analysis and provides the main sound characteristics of walking activity. For the second level of abstraction we provide higher-level acoustical features that better describe person identification.

PACS no. xx.xx.Nn, xx.xx.Nn

1. Introduction

Human gait carry valuable information which can be used for a broad range of applications, from person and gender identification, to diagnosis of Parkinson disease [1, 2, 3]. Gait recognition first appeared in computer vision with a view to provide human recognition at a distance [4]. For applications in indoors environments we can imagine that it can be cumbersome to equip a building with a 100% camera coverage system. This limitation does not apply for machine listening applications because of the very nature of sound propagation. As a result, there is a growing interest for indoors audio-based smart applications.

From an evolutionary perspective our ability to identify individuals using auditory cues is a crucial skill which may have survival value. For example, we have developed the ability to identify a person from his voice without any visual stimuli. Occupants' daily activities, like indoors mobility, provide us with rich auditory information. For example, we are listening to walking sounds while shopping in a mall, working in an open office, studying in the university library and much more. In all aforementioned places human-made sounds like speech and walking sound may dominate the social auditory scene. The major difference between speech and walking sound is that the latter do not have any obvious tonal components and may be the cause of impact sound transmision. In this study we are focusing on airborne sound that is captured via direct sound and its reflections. As a result, we do not see the term *walking sound* as equivalent to *footsteps sound*. For example walking sound might include the sound of clothes and shoelaces or even the sound of other accessories like a tinkling chain [5]. These are by-products of human walking activity yet they are indispensible part of walking sound.

In this experiment our aim was to identify individual persons based on airborne recordings of walking sounds. We conducted an indoors experiment in which four individuals were walking in a reverberant room equiped with a wooden floor. The signal analysis was based on magnitude analysis [6]. Our approach was to extract acoustical events of walking sound and calculate a characteristic data set of acoustic single quantity indicators. We call this a feature space. In order to identify individual walkers we then employed two fundamental techniques of statistical learning, principal component analysis (PCA) and linear discriminant analysis (LDA). Both aforementioned linear transformations can be used for dimensionality reduction of

(c) European Acoustics Association

Copyright © 2018 | EAA – HELINA | ISSN: 2226-5147
All rights reserved

a high dimensional dataset and for classification. The basic difference between PCA and LDA is that the former is unsupervised whereas the latter is a supervised classification technique. For example, PCA perfoms clustering which means that it is agnostic of class labels, whereas in LDA the class labels are already known. In our analysis the agnostic parameters were the higher-level characteristics of acoustical events and the known parameters the class labels of the individual walkers (ie. name of each person). The high-level acoustical characteristics refer to the synthesized acoustical features that we constructed using PCA.

The idea to treat acoustical events as events that are not necessarily related to footsteps is highly economical in the computational analysis. It also has great potential for applications with high ecological validity. For example in [7] the authors applied acoustic gait recognition on a staircase in order to have a predifined number of steps. In contrast our method is *footsteps agnostic*, and the amount of acoustical events does not necessarily has any relationship with the actual amount of footsteps.

2. Methods

2.1. Recording room and equipment

The recordings were performed at RISE, Research Institutes of Sweden. The recording room volume was $102.80m^3$ ($h = 3.41m, b = 4.92m, d = 6.13m$), and the wooden floor was Cross Laminated Timbre (CLT) with dimensions $4.0m \times 3.0m$ and thickness of $230mm$ (see Figure 1). Two seperate elements were installed in the lab each one of $1.5m$ width. The recording equipment was two B&K (half-inch free-field microphone, 6.3 Hz to 20 kHz, 200V polarization) and a dummy head (GRAS 45BB KEMAR Head & Torso). We used RME QuadMicII pre-amplifier, a Macbook Air (late 2010; El Capitan), and an EDIROL Hi-SPEED USB AudioCapture UA-101 soundcard (10IN/10OUT 24bit 192kHz). The recording software was REAPER v.5 digital audio workstation at 24bit/48kHz and block size 512 samples.

2.2. Participants and procedure

Four adults (one female; three males) were asked to follow an "hourglass" pattern and walk on the wooden floor about $1-2$ minutes. Using this pattern we were able to record walking sounds on the diagonal, and at the edges of the wooden floor. Each participant performed one walking session. One male participant was excluded from the analysis as an outlier due to the fact that his recording session had several spikes that exceeded the average magnitude levels. The cause of these spikes in the signal is unclear but it might be due to gravel sticked under the soles. Gravel is heavily used in northern countries as an anti-slip measure and

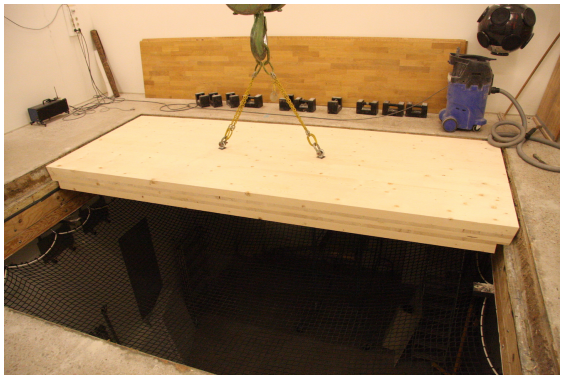

Figure 1. Wooden floor during installation in the lab.

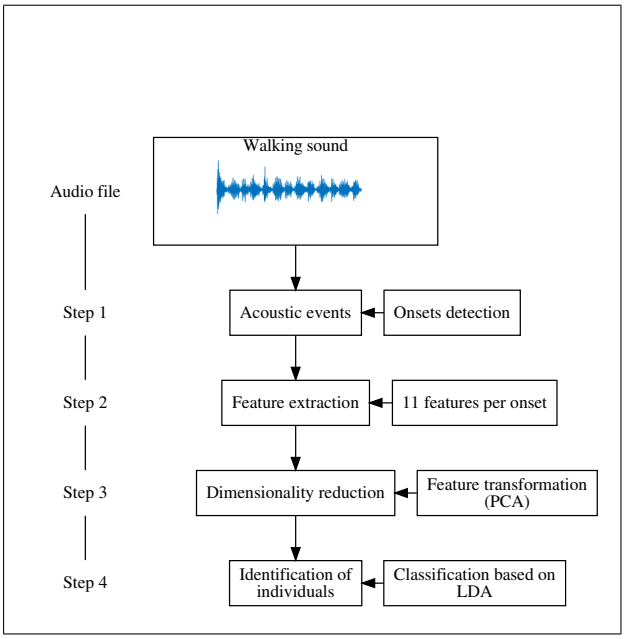

Figure 2. The four steps of computational analysis of walking sound.

a way to protect walking activities of civilians due to snow and ice weather conditions. The motivation to exclude this participant as an outlier was due the fact that we noticed high contributions in the variance of the extracted features. That become obvious in the principal component analysis.

2.3. Computational analysis

The computational analysis was based on magnitude analysis of the signal [6] and had four main parts as shown in Figure 2.

2.3.1. Acoustical events

Acoustic event detection of walking sounds was performed using SuperCollider programming environment using SCMIR library [8] for music information retrieval. The first step was to apply onsets detection in order to estimate the beggining of transient sounds [9]. Please note that the onsets do not necessary spot timestamps that indicate contact of the shoe pad with

the wooden floor. As we noted in the introduction we use the term *walking sound* more broadly than *footsteps*. Given this interpretation each onset might be trigged from different sound sources, like clothes, chains etc.

2.3.2. Feature extraction

The next step was to extract a set of 11 single quantity acoustical features for each onset. Specifically we extracted 11 acoustical features using SCMIR library. The corresponding class names in SuperCollider language are: `Loudness`, `SpectralEntropy`, `SpecCentroid`, `SpecPcile`, `SpecFlatness`, `FFTCrest`, `FFTSpread`, `FFTSlope` and `SensoryDissonance`. The list below provides a short descrption for each acoustical feature. For analytic descritpion of each class please see the online help file of SuperCollider.

1. **Loudness**: Variant of an MP3 perceptual model
2. **SpectralEntropy**: General peakiness of the spectral distribution
3. **SpecCentroid**: Spectral centroid, an indicator of perceptual brightness
4. **SpecPcile,0.99**: Cumulative distribution of the frequency spectrum, High values 0.95, 0.99 used for spectral roll-off
5. **SpecFlatness**: Has value of 0 for sine wave, 1 for white noise
6. **FFTCrest,2,50**: Spectral crest measure for the frequency range $2 - 50Hz$
7. **FFTCrest,50,500**: Spectral crest measure for the frequency range $50 - 500Hz$
8. **FFTCrest,500,8000**: Spectral crest measure for the frequency range $500 - 8000Hz$
9. **FFTSpread**: Measure of spectral spread (magnitude-weighted variance)
10. **FFTSlope**: Spectral slope, describes the reduction of spectral energy in high frequencies
11. **SensoryDissonance**: An indicator of perceptual roughness

The selection of these 11 features was based primarily on the available features of SCMIR library for SuperCollider. We discarded features that are related to tonality, like Chromagram and mel-frequency cepstral coefficients (MFCC), because walking sounds, as opposed to speech, do not have any tonal components. Furthermore, we didn't take into account any onsets statistics, like inter-onsets (IOI) mean and standard deviation. This is because people use to change their walking speed on different daily activities and walking speed has an effect on walking variability [10, 11].

Before the feature extraction we had mixed all four channels in REAPER DAW (2 low frequency range, 1 dummy head). On every mixed signal we applied a second-order low pass filter with cutoff frequency 9600Hz to reduce any contribution of high frequency noise that may cause high variations in PCA. We choose this high frequency cutoff because the spectrogram indicated walking activity up to 8000Hz. Then all audio signals we extracted a set of 11 acoustical features for every acoustical event (onset). We extracted 1357 onsets and 11 acoustical features per onset. The average recording time across all walking sessions was 123.67 seconds.

2.3.3. Dimensionality reduction based on PCA

From the two aforementioned steps (see Step 1 and Step 2 in Figure 2) we generated a feature space of acoustical features. This space has dimensions of $onsets \times features$ (1357×11). On this feature space we applied principal component analysis (PCA). This is a commonly used approach to reduce the high dimensionality of the data set. PCA is a linear transformation which creates a new synthetic feature subspace. The first synthetic dimension explains the largest percent of variance, the second principal component (ie. synthetic dimension) explains the second largest percent of variance and so on. Typical values of explained variance include 90%, 95% and 99%. Here we will select the principal components that explain the 90% of explained variance in order to generate the smallest possible feature subspace. On this subspace we applied varimax rotation which is a rotation of the orthogonal coordinate system that maximizes the variance. We will then attempt to provide an interpretation of these synthetic dimensions in order to identify which acoustical characteristics are better indicators of idiosyncratic walking sound.

2.3.4. Classification of individuals using LDA

The last step in the analysis was to perform linear discriminant analysis (LDA). The latter is one of the most fundamental classification techniques which is also used as an approach to reduce the total amount of dimensions. LDA uses quantities as predictors and predicts qualities, or nominal values. In our experiment our classes are the names of the four individuals. Whereas PCA does not assumes normal distributions, LDA is based on assumption of Gaussian distributions. On the other hand experimental evidence suggest that this assumption can be violated [12, 13].

That was the motivation to explore both normality and non-normality assumptions. For that purpose we examined skewness values of the 11 acoustical features and we assigned the absolute value of monad as upper and lower threshold. Based on this assumption we created a new feature space that has only five acoustical features out of the 11 extracted features. In the results section we will evaluate the performance of both feature spaces (with 5 and 11 features).

The skewness values for the 11 features are shown in the Table I below:

Table I. Skewness values for the set of 11 acoustical features.

Acoustical feature	*Skewness*
Loudness	1.366
SpectralEntropy	3.798
SpecCentroid	**0.727**
SpecPcile(0.99)	**0.241**
SpecFlatness	**0.544**
FFTCrest(2,50)	**0.517**
FFTCrest(50,500)	1.363
FFTCrest(500,8000)	1.160
FFTSpread	**0.244**
FFTSlope	-1.151
SensoryDissonance	3.638

3. Results

3.1. Transformation of feature space

Figure 3 shows the principal components loadings matrix for the 11 acoustical features that explain more than 90% of the variance. On the X axis is shown the percent of explained variance for each principal component (PC). On the Y axis is shown the set of all extracted acoustical features. This visual representation shows the contribution of each acoustical feature to every principal component [14]. For example, the first principal component (PC1) has major contributions from the features of *SpecPcile,0.99*, *SpecFlatness* and *FFTSpread*. The second has major contributions from *SpectralEntropy* and the third from *FFTCrest,2,50*.

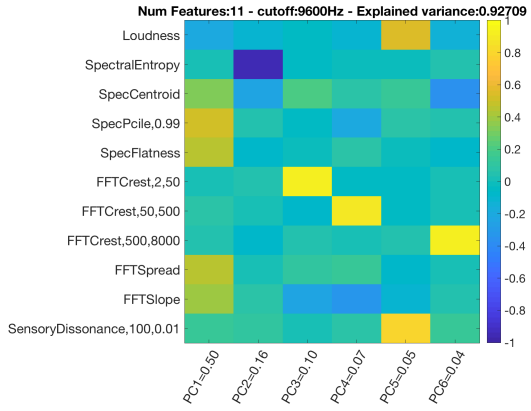

Figure 3. PC loadings matrix with 6 transformed features based on varimax rotation. The percent of explained variance is shown for each PC.

Figure 4 shows the principal components loadings matrix for the five dimensional feature space that explain more than 90% of the variance. We remind the reader that this feature subset was created with a view to approximate a normal distribution. We did that by selecting the features that have low skewness values. Whereas the two feature spaces cannot be compared in a quantitative manner, yet we see similarities. For example PC1 has major contribution from *SpectralCentroid*, *SpecFlatness* and *FFTSpread*. All these features are present in PC1 of the PC loadings matrix for the 11 acoustical features (see Figure 3). PC2 and PC3 are better described by *FFTCrest,2,50* and *SpecPile,0.99* respectively. As a result, we see that all aforementioned acoustical features are present in the first three principal components in Figure 3. Following this intuition we may assume that the 3D feature subspace preserves the most important acoustic information of the 6D feature subspace. This is because the first three PCs of the 6D feature subspace explain more than 75% of variance of the 11D feature space.

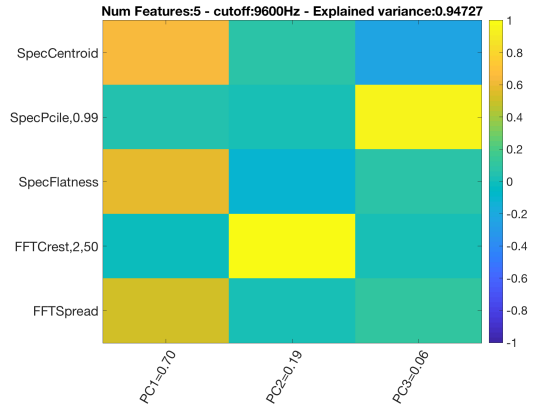

Figure 4. PC loadings matrix with 3 transformed features based on varimax rotation

3.2. Prediction of class labels

We build a classifier based on LDA to evaluate the performance of both 3D and 6D the feature subspaces that we synthesized based on the 5D and 11D feature spaces. The LDA classifiers were build using 10-fold crossvalidation on the 85% of the data set and we estimated the misclassification error. Then we used these classifiers to predict an unseen data set for the remaining 15% of the data. The classification error of our model for the 3D subspace was 22.81% and the classification performance of the unseen data set was 73.53%. On the other hand, for the 6D subspace the classication error was 25.15% and the classification performance on the unseen data set was 53.43%. Figure 5 shows the projection of the unseen data set (15%) on the linear discriminants.

3.2.1. Linear discriminant loadings

Figure 6 shows the linear discriminants loadings matrix based on varimax rotation. This visual representation, similar to the principal components loadings

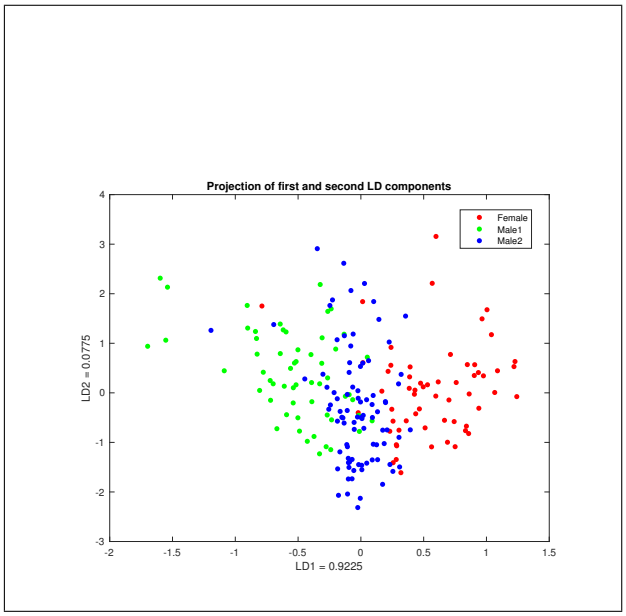

Figure 5. Projection of the unseen data set (15%) on the linear discriminants.

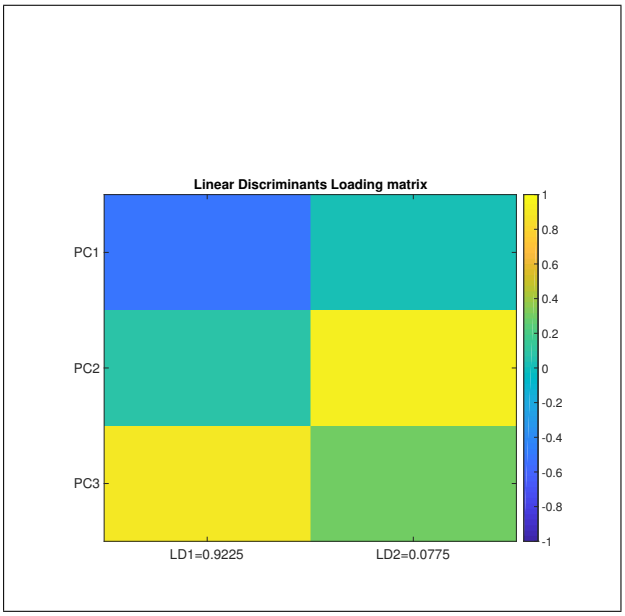

Figure 6. Linear Discriminant loadings matrix based on varimax rotation.

matrix, shows the contribution of PC1, PC2 and PC3 to the linear discriminants (LD1 and LD2). This is the next level of abstraction that corresponds to a newly synthesized feature subspace that has only two components. The first discriminant corresponds to 92.25% percent and the second discriminant to 7.75% of the acoustic information that is carried withing the 3D feature subspace of PC1, PC2 and PC3. We see that the first linear discriminant is negatively correlated with PC1 and positively correlated with PC3. PC1 is a combination of perceptual brightness, spectral flatness and spectral spread. The second linear discriminant is dominated by PC2 which has major contributions from low frequency components.

4. Discussion

The comparison of the two synthesized feature subspaces that we created based on PCA showed that the 3D feature subspace is better predictor for identification of individuals. Following that linear transformation we provided a visual representation that shows the contribution of this feature subspace to the linear discriminants (see Figure 6).

Whereas previous studies in gait recognition have been focused on classification performance our main goal was to employ fundamental techniques that afford high interpretability. Our motivation was to present a methodogy that might be useful for applications across a broad range of applications. In the context of building acoustics, sound sources related to human activities may vary to a great extend. In our methodology we demonstrate a manner that enable us to construct high-level acoustical features that explain the acoustical variation of walking sound. We provide two levels of abstraction for the high-level acoustical features. The first abstraction corresponds to the principal components loadings matrices (see Figure 3 and Figure 4) which provides an excellent visual representation of higher level acoustical indicators of walking sound. The second level of abstraction corresponds to higher-level acoustical features that better discriminate individuals. Figure 6 shows the contribution of the principal components to the linear discriminants. For example, that means that whereas PC2 has major contribution in the variance of walking sounds characteristics it does not play an important role in person identification. We see that PC1 of the 3D feature subspace is better described by perceptual brightness, spectral flatness and spectral spread. We can possibly interpret this component as “spectral richness”. Ultimately, we see that this view is enhanced by the negative correlations that PC1 exhibits with PC3 in LD1 as we can interpret the latter as squeaking sounds. These are sounds that are produced when we are doing manoeuvres, like turning around, stop walking or other highly idiosyncratic walking sounds.

The advantage of using linear transformations to perform identification of walkers is that PCA and LDA are “transparent” techniques that are easy to interpret. For example, in [15, 16, 17] the classification techniques are based on HMM, nearest neighbour (NN) and hierarchical clustering. In HMM and NN the interpretation is not straight forward, and in hierarchical clustering the interpretability can be quite difficult. This is known as trade-off between prediction and interpretability. That means that we may be able to develop a better prediction method using HMM, NN or hierarchical clustering, yet we cannot

seak for a straight forward interpretation of our predictive model.

Acknowledgement

This project has been funded by ACOUTECT Innovative Training Network.

References

[1] DeLoney, C. (2008). Person identification and gender recognition from footstep sound using modulation analysis

[2] Lee, H., Guan, L., & Lee, I. (2008). Video analysis of human gait and posture to determine neurological disorders. EURASIP Journal on Image and Video Processing, 2008(1), 380867.

[3] Wang, L., Tan, T., Ning, H., & Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. IEEE transactions on pattern analysis and machine intelligence, 25(12), 1505-1518.

[4] Nixon, M. S., Tan, T., & Chellappa, R. (2010). Human identification based on gait (Vol. 4). Springer Science & Business Media.

[5] Makela, K., Hakulinen, J., & Turunen, M. (2003). The use of walking sounds in supporting awareness. Georgia Institute of Technology.

[6] Vigran, T. E. (2014). Building acoustics. CRC Press.

[7] Alpert, D. T., & Allen, M. (2010, September). Acoustic gait recognition on a staircase. In World Automation Congress (WAC), 2010 (pp. 1-6). IEEE.

[8] Collins, N. (2011). SCMIR: A SuperCollider music information retrieval library. In ICMC.

[9] Stowell, D., & Plumbley, M. (2007, August). Adaptive whitening for improved real-time audio onset detection. In Proceedings of the 2007 International Computer Music Conference, ICMC 2007 (pp. 312-319).

[10] Hausdorff, J. M. (2007). Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking. Human movement science, 26(4), 555-589.

[11] Jordan, K., Challis, J. H., & Newell, K. M. (2007). Walking speed influences on gait cycle variability. Gait & posture, 26(1), 128-134.

[12] Duda, R. O., Hart, P. E., & Stork, D. G. (1973). Pattern classification (Vol. 2). New York: Wiley.

[13] Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation. Knowledge and information systems, 10(4), 453-472.

[14] Alluri, V., Toiviainen, P., Jäskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage, 59(4), 3677-3689.

[15] Geiger, J. T., Kneißl, M., Schuller, B. W., & Rigoll, G. (2014, November). Acoustic gait-based person identification using hidden Markov models. In Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop (pp. 25-30). ACM.

[16] Pan, S., Wang, N., Qian, Y., Velibeyoglu, I., Noh, H. Y., & Zhang, P. (2015, February). Indoor person identification through footstep induced structural vibration. In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (pp. 81-86). ACM.

[17] Orr, R. J., & Abowd, G. D. (2000, April). The smart floor: A mechanism for natural user identification and tracking. In CHI'00 extended abstracts on Human factors in computing systems (pp. 275-276). ACM.